

## Crítica del Libro "Ética para Máquinas", de José Ignacio Latorre, Ariel, Barcelona, 2019, 320 págs.

MARC-ABRAHAM PUIG HERNÁNDEZ

*Universidad de Barcelona*

*marcph@icab.cat*

### RESUMEN

Actualmente predomina la acción de los seres humanos en relación con las máquinas. Latorre propone, partiendo de un recuento de la evolución de las máquinas, un escenario ficticio donde las máquinas han superado al ser humano especialmente, en su actividad neuronal y de pensamiento. El autor discurre sobre cuestiones de moral humana aplicada en máquinas, a fin de llevar a una reflexión respecto del avance de la tecnología, la inteligencia artificial y la relación de los humanos con este medio.

### ABSTRACT

Nowadays, human action prevails over their relation with machines. Latorre, from reminding machines evolution, proposes a fictional scenario where machines have overcome humans, especially in neuronal activity and thinking. This author reasons on human moral questions applied to machines, in order to bring into a consideration on technology and artificial intelligence advance, and human relationship with this environment.

**PALABRAS CLAVE:** Ética, física, computación, moral, inteligencia artificial

**KEYWORDS:** Ethics, Physics, Computing, moral, artificial intelligence



Copyright© Marc-Abraham Puig Hernández

Se permite el uso, copia y distribución de este artículo si se hace de manera literal y completa (incluidas las referencias a i-Latina), sin fines comerciales y se respeta al autor adjuntando esta nota. El texto completo de esta licencia está disponible en: <http://creativecommons.org/licenses/by-nc-sa/2.5/es/legalcode.es>

Que un catedrático de física teórica se moleste en compartir sus reflexiones sobre cualquier aspecto con cierta relevancia para la sociedad siempre es motivo de celebración. En este caso, sobre los aspectos convenientes, y también los peligrosos, relativos a nuestra relación con las máquinas, incluso de aquellas que podemos denominar inteligentes y que tarde o temprano nos dejarán atrás. Un físico escribiendo sobre ética constituye una carta de presentación suficientemente atractiva para recomendar la lectura del libro.

Cierto es que no está escrito con el vocabulario o la incisión analítica propia del debate académico, pues es una obra sin barreras para el público, expuesta con sencilla claridad y haciendo accesibles conceptos de diversos campos, como los propios de la física, la computación o la propia filosofía. Valga decir que el acervo científico del autor se traslada a la obra con una exposición a menudo más cercana a la lógica proposicional que al derroche discursivo característico de las humanidades. Pero de ningún modo esta cuestión estilística es un impedimento para disfrutar de la obra; la intención del autor no es otra que invitar a la reflexión sobre el avance tecnológico, la relación de los humanos con las máquinas y la inteligencia artificial. Qué menos que recoger el guante y aceptar esa invitación.

Ética para máquinas está estructurado como un sendero que une una evolución histórica, a partir de la que el autor nos va introduciendo la relación del ser humano con las máquinas y la tecnología, la constante evolución de éstas y la perenne evaluación del rendimiento bueno y malo del uso al que se han destinado, con un escenario hipotético, o de ciencia ficción, si se prefiere, cuando las máquinas intelectualmente superiores a nosotros sean una realidad. A menudo, como bien queda reflejado en las líneas de la obra, la reflexión ha llegado muy tarde, una vez se han producido las consecuencias. Ese es el principal motivo que lleva al autor a compartir este trabajo: tomar una pausa y prestar atención a las posibles consecuencias derivadas de los cambios que se producirán con la irrupción de las nuevas tecnologías, especialmente con las máquinas inteligentes que superarán al ser humano en todo. Ello justifica una primera parte de observación histórica y de

evaluación del rendimiento bueno o malo y una segunda de apoyo en las lecciones previas para no cometer los mismos fallos en un futuro próximo.

Para nuestra sorpresa, la postura de LATORRE es optimista. Defiende que las consecuencias de máquinas cuya inteligencia nos relegue serán positivas, aunque no tenga que ser necesariamente a corto plazo, pues uno de los primeros problemas que deberemos hacer frente será el de la eliminación de muchos puestos de trabajo. Nos advierte acerca de que la implantación de esta nueva tecnología traerá consigo que diferentes generaciones compitan por un mismo puesto de trabajo. Por ejemplo, se pelearán abuelos, hijos y nietos por realizar la misma labor remunerada. Parece ser, según el autor, que esta lucha vertical, generacionalmente hablando, vendrá a invertir la clásica polarización izquierda-derecha, obrero-empresario, mano de obra-oferta de trabajo, que ha caracterizado el problema del desempleo durante tanto tiempo y especialmente agudizado en las crisis del siglo XX y comienzo del XXI. En mi opinión, más que invertir los polos del desempleo, ambos tipos configurarán los ejes de las coordenadas sobre las que deberemos aprehender las dificultades de un desempleo caracterizado por que las máquinas inteligentes asuman funciones cualificadas. En otras palabras, al no hacer falta ni experiencia ni especialización, que ahora asumirán las máquinas, las generaciones competirán entre sí.

Si bien podemos afirmar, por un lado, que desde la Revolución Industrial el progreso tecnológico siempre ha traído un incremento de las colas del desempleo y por ello un nuevo avance no debería suponer un cambio radical de escenario sino otra vuelta de tuerca al caso, por otro lado, las tecnologías capaces de aprender faenas intelectuales convierten en un sinsentido la constante especialización del trabajo, tan propio de nuestros días. Una máquina hará esas tareas mejor que un ser humano y, en principio, no habrá buenas razones, o mejores razones, para que los empleadores opten por esos trabajadores sin alma tan eficientes y tan poco quejicas. Pero me apunto a la filosofía optimista de nuestro autor. Tal vez la parte buena de este embrollo será que así dispondremos de más tiempo para dedicarnos a una vida contemplativa. Tal vez aumente la demanda en las carreras de humanidades. Tal vez,

eso sí, si somos capaces de apagar nuestro último Smartphone durante un rato y dedicar nuestro tiempo a este tipo de quehaceres en lugar de consumir ocio compulsivamente, lo cual vendrá personalizado por el trabajo de las nuevas tecnologías y será realmente difícil.

No hemos dejado constancia acerca de por qué sostiene el autor que las máquinas inteligentes nos superarán. El entrenamiento de redes neuronales artificiales ha progresado más allá del punto de inflexión que suponía para su aprendizaje que el ser humano fuese una pieza necesaria para continuar mejorando. Es decir, para seguir aprendiendo, las redes neuronales artificiales compiten una con otra. Redes adversarias se entrenan una a otra a partir del modelo de acierto y error. Si una máquina consigue inducir a su adversaria en el error, se descarta esa respuesta y se insiste hasta lograr perfeccionar la respuesta. De esta manera, el aprendizaje de estas redes es cada vez mayor, más rápido y refinado y con mayor número de aciertos. En otras palabras, para continuar mejorando, las máquinas inteligentes necesitan competir con máquinas igual de inteligentes una vez han superado el entrenamiento, o el aprendizaje, al que las sometían los humanos.

Es en este sentido que laboralmente jamás podremos competir con máquinas entrenadas para desempeñar un determinado trabajo. Por un lado, uno se acuerda del antipático funcionario judicial que semana tras semana pone alguna pega para la tramitación de los escritos o la notificación de providencias. Piensa que por fin una máquina inteligente será mucho más eficiente para agilizar la Administración de Justicia. Pero, por otro lado, abruma pensar que ese funcionario judicial, quien ya no podrá dedicarse a ello porque le sustituirá una máquina inteligente, ahora se verá reconvertido en un abogado o en un procurador más en la ya concurrida circunscripción de Barcelona. Ante este panorama ya no sabe tan mal esperar cinco minutos para que tramite el escrito si de esta manera no escasea el trabajo, puesto que evitaríamos que nuevos letrados tuvieran que pelearse con otros cientos de nuevos competidores que llevan a cabo el mismo trabajo. Si las máquinas inteligentes se destinaran a ayudar a que el funcionario desempeñara mejor sus funciones, a que

no se entretuviera con el farragoso papeleo y que finalmente pudiera expedir los documentos ipso facto, ya supondrían una mejora sustancial de la Justicia. Nuestro autor afirma que estamos condenados a enterrar aquellos trabajos que las máquinas inteligentes puedan realizar mejor y con mayor eficacia que un ser humano; sin embargo, no puedo más que tratar de buscar una aplicación a esas nuevas tecnologías que bien nos eviten realizar el trabajo mecánico, incluso el intelectualmente mecánico, sin tener que suprimir los actuales puestos. Puede que mi postura sea ingenua, que la suerte esté echada y que ya hayamos condenado ese tipo de oficios. Ahora bien, el uso público de las máquinas inteligentes está condicionado por políticas y por la adecuación de las mismas a las exigencias del conjunto del ordenamiento jurídico. Cabe la posibilidad de que acompañemos a la implantación de máquinas inteligentes con la supervisión del ser humano, o simplemente que la última palabra siempre dependa de la intervención de la persona. Y como esto último se eleva por convención, en mi opinión, aún no está todo dicho.

Tras el paseo histórico, el planteamiento de escenarios como la reducción de puestos de trabajo y el repaso de un buen número de pensadores, la segunda mitad del libro plantea cuestiones más difíciles de asir. Entre ellas, encontramos el planteamiento de la relación con inteligencias superiores a la nuestra, pues una vez dejaremos de ser imprescindibles para que continúe avanzando el aprendizaje las máquinas caminan hacia la singularidad (me hubiera gustado encontrar alguna referencia sobre lo bueno y lo malo de que el ser humano deje de ser el centro del mundo intelectual, pues supone un verdadero manotazo a nuestro egocentrismo en tanto que especie; mejor dicho, el egocentrismo de cómo el ser humano se ve a sí mismo como especie. No obstante, esta cuestión se sitúa más allá de lo que pretende presentar el autor), o la gran pregunta de si más allá de vernos superados intelectualmente puede llegar a imitarse el alma humana en una máquina o una inteligencia artificial.

Ciertamente, el tema más espinoso que aborda José Ignacio LATORRE no es la inteligencia artificial en los términos que la define ni el paso a especular en un

escenario propio de la ciencia ficción, sino el del alma. Creamos o no en el alma, ya de por sí difícil una temática difícil de abordar y definirla, estamos ante una nota característica del ser humano. La posibilidad de que ésta sea imitada en una máquina o una inteligencia artificial nos debería hacer reflexionar sobre si el derecho debe considerar que estamos ante un sujeto autónomo: ¿estaríamos ante un sujeto autónomo en los mismos términos en que el derecho reconoce a una persona? Según el autor, llega el momento en que se hace evidente la necesidad de repensar el legado moral que vamos a transmitir a inteligencias superiores a la humana, que serán capaces de sentir emociones y reflejar sentimientos.

Desde luego yo no puedo estar completamente seguro de la respuesta a la anterior interrogante (porque parece difícil encontrar consenso acerca del papel que juegan las emociones en el debate moral, y aunque no entraré aquí en este debate, simplemente asumiré que existe algún tipo de relación). Coincido con LATORRE en que nunca ha habido unanimidad para responder a la pregunta “qué es el alma”. Realmente es una tarea ardua. Ahora bien, el autor insiste en la idea de transmitir emociones y sentimientos a una inteligencia artificial. Esta transmisión es la que dotaría a una máquina inteligente del contenido de lo que supuestamente estaría compuesta nuestra alma humana. Si realmente podemos legarles emociones, sentimientos e incluso invitarlas al debate moral, el autor se acerca mucho a la quintaesencia del asunto con la transmisión de todo lo anterior. Ahora bien, seguramente podríamos afinar un poco más.

Retrocedamos un paso hasta el instante anterior de afirmar que nuestras emociones y sentimientos residen en el alma. No pretendo discutirlo, sino tratar de coger un desvío previo hacia otro sendero de reflexión. Ciñámonos por un instante al alma como ese algo inmaterial que acompaña a nuestro cuerpo; ese impulso que nos lleva a vivir la vida con sentido. Y nada más. Todo lo que viene más allá de esta característica es realmente difícil de acotar: cuántas emociones existen, la relación entre sentimientos, la experiencia de la vida necesaria para acertar en las decisiones morales... Lo primero no sería incompatible ni con la propuesta del autor ni con

ninguna de las referencias de la historia del pensamiento que nos ofrece, pues en el impulso de vivir la vida con sentido encontramos la libertad, los sentimientos y otras muchas características humanas que nos permiten precisar, según nosotros mismos, cuál es ese sentido, y buscar las respuestas a preguntas sobre nuestra existencia, nuestro paso por la vida, que tan bien circunscribiríamos en los márgenes intangibles del alma.

Más que preguntarnos por el legado emocional, sentimental y moral que dejaríamos a las máquinas que nos superarán, deberíamos plantearnos si dichas inteligencias artificiales serían capaces de tener el impulso que les lleve a dotar de algún sentido su propia existencia. Porque lo que queda claro de la exposición del autor (y digo claro porque en todo momento nos acompaña, en tanto que lectores, planteándonos una serie de preguntas sin asumir promesas de futuras respuestas que luego no se cumplan) es que nosotros somos los que programamos un algoritmo con sentimientos, emociones, con capacidad de simularlos e incluso con algún código moral. Y si esto es así, nosotros seríamos el impulso de esas entidades y de alguna manera estableceríamos previamente el sentido de su existencia.

Si somos nosotros quienes programamos el alma de las máquinas, con cierta legitimidad podemos afirmar que el alma de dichas inteligencias también será artificial. A no ser que consideremos que el alma es propiamente algo relacionado íntimamente con el ser y que ese impulso no puede serle algo ajeno al ser humano o al ser artificial. Si consideramos esto último una condición necesaria del ser, nuestra relación con las máquinas siempre se medirá en los parámetros sujeto-objeto. Por el contrario, desde el momento en que identificamos que cualquier ser que tenga el impulso de buscar el sentido de su existencia merece nuestro reconocimiento, con independencia del origen de ese impulso, estaríamos ya hablando de máquinas e inteligencias entendidas como agentes morales autónomos.

El autor, al final de la obra, nos abre la posibilidad de pensar en ello a través de la libertad como contenido factible de ser legado a una inteligencia artificial.

Desde luego, tras haber elevado el nivel de abstracción ahora chirría un tanto pensar en una inteligencia artificial como ser autónomo, programada para que su impulso por dotar de sentido a su existencia se asemeje al de un ser humano, pues cuesta imaginar un ejemplo más humano que el propio humano. Pero en estos casos podemos ayudarnos de la ciencia ficción, como el momento final de Blade Runner en que el replicante Roy Batty, un ser inteligente creado a imagen y semejanza del ser humano pero que lo mejora en todos los aspectos, salva la vida de su perseguidor y cuando está a punto de morir (si es que podemos emplear el término morir para un ser artificial, a no ser que reconozcamos el exitus vitae de algo parecido a su espíritu y otra vez topamos de alguna manera con el alma), realiza un discurso de alto nivel moral, cuyo sentido ni siquiera llega a ser planteado, en muchos casos, por los propios seres humanos. En efecto, un ejemplo de algo más humano que el propio ser humano.

Llegados a este punto, interesa destacar los tres niveles de ética artificial que explica LATORRE. Una red neuronal artificial es capaz de aprender. En la arquitectura de la red se codifica la información necesaria para que actúe, a saber, se fijan las reglas necesarias para que la inteligencia artificial ofrezca respuestas. Un de estas reglas es la función error, la que codifica el criterio que debe aprender la red. La ética práctica de una red neuronal artificial se codificaría como una función error, y como decíamos, ahí el autor identifica tres niveles. El primero, la función error elemental, que mide el éxito de una red según minimiza el margen de error en funciones objetivables, como un algoritmo para predecir el tiempo. El segundo, la función error compleja, que combina diferentes funciones error elementales, tal y como pasa en un coche autónomo al mezclar funciones de parámetros visuales, sonoros, datos de tráfico... Y el tercero, el de las funciones error automejoradas, que permite a la propia máquina inteligente modificar las funciones error iniciales. La pregunta sería si lo que hemos definido como alma artificial junto con el legado moral y emotivo que programaremos en estas máquinas puede constituir una de esas funciones error alterables por la propia inteligencia artificial o si, por el contrario, lo



que puede no y no puede alterar la máquina también está codificado por el programador.

La primera opción sería muy tentadora a efectos de investigaciones sobre los elementos que inciden en la conducta humana. Esto es, saber cómo evolucionan los sentimientos tras desarrollar ciertas emociones y qué tipo de debate moral abre la propia máquina al respecto. Este tipo de experimentos servirían para la regulación de conductas en casos difíciles en los que no está claro qué criterio debe ser el determinante para tomar una decisión.

La segunda nos situaría ante el código ético del profesional y el tipo de buenas funciones (si es que se me permite usar esta expresión) que codifica en la red neuronal. En otras palabras, siempre caería la sospecha sobre la incidencia humana en la programación, en el acotamiento de los términos en que puede desarrollarse emocional y moralmente una inteligencia artificial.

En conclusión, estamos ante un agradable libro de ética, si es que por ética entendemos el arte de vivir bien y que gracias a la experiencia podemos reflexionar para intentar acertar en nuestras decisiones, en este caso, acerca de nuestra relación con las máquinas inteligentes. Seguramente, como se dijo más arriba, el lector erudito en filosofía moral identificará la ausencia de algún parámetro formal o material. Pero no es esa la finalidad del libro, sino la de compartir una reflexión sobre asuntos que nos envuelven a todos; un intento de acertar en la decisión sobre el uso que daremos a esta tecnología. Una obra que nos ha servido para identificar cuestiones relevantes para la filosofía del derecho, de las que hemos querido compartir las relativas a la destrucción de los puestos de trabajo y a los elementos que pueden configurar la relación sujeto humano y sujeto máquina. Porque, ante un agente autónomo que nos mejore intelectualmente y asimismo perfeccione nuestra parte humana, aunque su origen fuese artificial reconoceríamos que está mejor condición que nosotros mismos para crear buenas leyes, ¿o tal vez no?